REMARKS

This Amendment responds to the Office Action dated February 10, 2005 in which the Examiner objected to claim 2, rejected claims 1-4, 9-13 and 16-20 under 35 U.S.C. §102(a) and rejected claims 5-8 and 14-15 under 35 U.S.C. §103.

As indicated above, a typographical error in claim 2 has been corrected. Therefore, Applicant respectfully requests the Examiner approves the correction and withdraws the objection to claim 2.

Claim 1 claims a method and claim 13 claims a system for automatically finding one or more answers to a natural language question in a computer stored natural language text database. The natural language text database has been analyzed with respect to syntactic functions of constituents, lexical meaning of word tokens, and clause boundaries. The natural language question comprises a question clause. The method comprises the steps of: first, analyzing a computer readable representation of the question clause with respect to syntactic functions of its constituents and the lexical meaning of its word tokens. In response to the analysis step, a set of conditions for a clause in the natural language text database is defined to constitute an answer to the question clause. The conditions comprise a condition stipulating that, for a clause in the natural language text database to constitute an answer to the questions clause, at least one of the constituents of the question clause should have a corresponding constituent in the clause having the same syntactic function and an equivalent lexical meaning. Clauses are identified in the natural language text database that satisfy the conditions. Answers to the question clause are returned by means of the identified clauses that match the conditions.

Through the method and structure of the claimed invention defining a set of conditions for a clause in the natural language text database to constitute an answer to a question clause where the conditions comprise a condition stipulating that, for a clause in the natural language text database to constitute an answer to the questions clause, at least one of the constituents of the question clause should have a corresponding constituent in the clause having the same syntactic function and an equivalent lexical meaning, as claimed in claims 1 and 13, the claimed invention provides a method of automatically finding one or more answers to a natural language question that are not domain specific and that deliver answers to questions with high precision. The prior art does not show, teach or suggest the invention as claimed in claims 1 and 13.

Claim 16 claims a method of automatically finding one or more answers to a natural language question in a computer stored natural language text database. The natural language text database has been analyzed with respect to clause boundaries, constituents, syntactic functions of the constituents, word tokens of the constituents and lemmas of the word tokens. The natural language question comprises a question clause having an interrogative pronoun. The method comprises the steps of: first, analyzing a computer readable presentation of the question clause with respect to constituents, syntactic functions of the constituents, word tokens of the constituents and lemmas of the word tokens. The interrogative pronoun is identified in the computer readable presentation of the question clause. The syntactic function of interest is determined based on the analyzing step and the identified interrogative pronoun. A constituent having the determined syntactic function in the question clause is identified. A word token is selected which is a

lexical head of the identified constituent in the question clause. A set of conditions is defined and comprises a condition stipulating that, for a clause in the natural language text database to constitute an answer to the question clause, the clause comprises a word token having the same lemma as the selected word token of the question clause and is comprised in a constituent having the same syntactic function as the identified constituent in the question clause. Clauses in the natural language text database are identified that satisfy the conditions. Finally, answers to the question clause are returned by means of the identified clauses that satisfy the conditions.

Through the method of the claimed invention a) determining a syntactic function of interest based on analyzing step and an identified interrogative pronoun, b) identifying a constituent having the predetermined syntactic function in a question clause and c) defining a set of conditions comprising a condition stipulating that, for a clause in the natural language text database to constitute an answer to the question clause, the clause comprises a word token having the same lemma as the selected work token of the question clause and is comprised in a constituent having the same syntactic function as the identified constituent in the question clause, as claimed in claim 16, the claimed invention provides a method of automatically finding one or more answers to a natural language question that are not domain specific and that deliver answers to questions with high precision. The prior art does not show, teach or suggest the invention as claimed in claim 16.

Claims 1-4, 9-13 and 16-20 were rejected under 35 U.S.C. §102(a) as being anticipated by *Wang et al.* ("A Question Answering System Developed as a Project in a Natural Language Processing Course").

Applicant respectfully traverses the Examiner's rejection of the claims under 35 U.S.C. §102(a). The claims have been reviewed in light of the Office Action, and for reasons which will be set forth below, Applicant respectfully requests the Examiner withdraws the rejection to the claims and allows the claims to issue.

Wang et al. appears to disclose the overall architecture of a system as depicted in Figure 1. The story sentences and its five questions (who, what, where, when, and why) are first preprocessed and tagged by the Brill part-of-speech (POS) tagger distributed with the Deep Read system. This tagged text is then passed to the Name Identification Module, which updates the tags of named entities with semantic information and gender when appropriate. The Partial Parser Module then takes this updated text and breaks it into phrases while attempting to lexically disambiguate the text. The Pronoun Resolution Module is consulted by the parser in order to resolve pronouns before passing partially parsed sentences and questions to the Sentence-to-Question Comparison Module. The Comparison Module determines how strongly the phrases of a sentence are related to those of a question, and this information is passed to several modules which attempt to learn which features of the comparison are the most important for identifying whether a sentence is a strong answer candidate. (pages 28-29, section 2) The Name Identification Module expects as input a file that has been tagged by the Brill tagger distributed with the Deep Read system. The most important named entities in the Remedia corpus are the names of people and the names of places. (page 28, section 2.1) The Partial Parser Module follows sequentially after the Name Identification Module. The input is the set of story sentences and questions, such that the words in each are tagged with POS tags and the names are marked with

type and gender information. (page 29, section 2.2) There were two methods we used to construct the lexicon: open lexicon, which includes all words from the development set along with all determiners, pronouns, prepositions, particles, and conjunctions (these words are essential to achieving good sentence segmentation), and closed lexicon, which includes all of the development and testing words. Each word entry contains information about its root word (if there is one), its lexical category (or categories) along with a corresponding set of allowable features and their corresponding set of allowable features and their corresponding values. Lexical categories include noun, verb, pronoun, propernoun, adjective, adverb, preposition, particle, conjunction, determiner, cardinal, ordinal, predetermined, noun modifier, and mouth. Feature types used in the lexicon include subcat, gender, agr, case, vtype (e.g., progressive), mood, gap, inverted, voice, behavior (e.g., mass), type (e.g., interrogative, relative), SemType, and conjtype (e.g., noun-type, verb-type, etc.) We hypothesized that SemType should play a significant role in improving question answering performance, but the choice of semantic granularity is a difficult problem. (page 30, section 2.2.1) The parser segments each sentence into either a noun phrase (NP), a verb phrase (VP), or a prepositional phrase (PP), each with various feature sets. (page 30, section 2.2.2.) A pronoun resolution module was developed using the rules given in Allen's text (Allen, 1995) along with other rules described in the work of Hobbs (Hobbs 1979). The module takes as input the feature-augmented and segmented text provided by the partial parser. Hence, the words are marked with lexical (including gender) and semantic feature information, and the phrase structure is also available. After the input file is provided by the Partial Parser Module, the Pronoun Resolution Module searches for the pronouns by

looking through the NPs identified by the partial parser. Candidate antecedents are identified and a comparison of the features is made between the pronoun and the possible antecedent. The phrase that passes the most rule filters is chosen as the antecedent. (page 31, section 2.3) The Sentence-to-Question Comparison Module takes as input a set of tagged stories, for which phrase types and features have been identified. The semantic and syntactic information is coded as shown in Figure 2 (using XML tags). A mechanism to quantify a qualitative comparison of questions and sentences has been developed. The comparison A comparison is made using phrase-to-phrase comparisons between each sentence and each question in a passage. In particular, NP-to-NP, VP-to-VP,PP-to-PP,and NP-to-PP comparisons are made between each sentence and each of the ve questions. These comparisons are stored for each sentence in the following arrays. Note that in these arrays Q varies from 1 to 5, signifying the question that the sentence matches. F varies over the features for the phrase match. Values for these comparison matrices were calculated for each sentence by comparing the features of each phrase type in the sentence to features of the indicated phrase types in each of the five questions. The individual matrix values describe the comparison of the best match between a sentence and a question for NP-to-NP (the three feature match scores for the best matching NP pair of the sentence and question Q are stored in CNQ]), VP-to-VP (stored in CVQ]), PP-to-PP (stored in CPQ]), and PP-to- NP (store in CPNQ]). Selecting the phrase comparison vector for a phrase type that best matches a sentence phrase to a question phrase was chosen as a heuristic to avoid placing more importance on a sentence only because it contains more information. Comparisons between features were calculated using the following equations. The

first is used when comparing features such as Base, NeedBase, and Prep, where a partial match must be quantified. The second is used when comparing features such as SemType, AGR, and Tense where only exact matches make sense. The matrices for the development set were provided to the algorithms in the Answer Module for training the component answer classifiers. The matrices for the testing set were also passed to the algorithms for testing. Additionally, specific information about the feature values for each sentence was passed to the Answer Module. (pages 31-32, section 2.4)

Thus, *Wang et al.* is directed to improving the accuracy of question answering. However, *Wang et al.* does not show, teach or suggest using the syntactic function in the actual identification of answers to questions. Applicant notes that *Wang et al.* identifies syntactic functions in an initial analysis, but does not use it in the actual matching criteria as indicated in the table in section 2.4 page 32. Thus, *Wang et al.* teaches away from the claimed invention as claimed in claims 1, 13 and 16 as explained in more detail below.

Wang et al. discloses using a SemType as playing a significant role in improving question answering performance (section 2.2.1 second paragraph, page 30.) Thus, nothing in *Wang et al.* shows, teaches or suggests a) defining a set of conditions, where the conditions comprise a condition stipulating that, for a clause in the natural language text data base to constitute an answer to a question clause, at least one of the constituents of the question clause should have a corresponding constituent in the clause having the same syntactic function and an equivalent lexical meaning as claimed in claims 1 and 13 or b) defining a set of conditions comprising a condition stipulating that, for a clause in the natural language text database to

constitute an answer to the question clause, the clause comprises a word token having the same lemma as the selected word token in the question clause and being comprised in a constituent having the same syntactic function as the identified constituent in the question as claimed in claim 16. Rather, *Wang et al.* discloses that the text is parsed and the words tagged with a semantic function (LABEL) but *Wang et al.* does not disclose the use of the LABEL of each word directly in the matching rules for finding an answer to a question (see the table in section 2.4, page 32). Instead, *Wang et al.* merely discloses a semitype plays a significant role.

The method as claimed in claim 16 is directed to a method where the syntactic function and lemma of word tokens play a significant role together with the use of information carried by the interrogative pronoun. Thus, the method as claimed in claim 16 does not rely on semantic types as in *Wang et al.* but relies upon syntactic functions of constituents. In particular, the method of claim 16 determines a syntactic function of interest in view of the interrogative pronoun and not a semantic type as in *Wang et al.* Furthermore, matching criteria according to claim 16 is the same syntactic function between constituent in potential answers and determine constituents in question. This increases the precision of the method as a word having the same semantic type and two different clauses may well have different syntactic functions in these clauses. For example, if you were looking for the answer to the question "who the dog bit?", you are generally not interested in receiving answers also to "who bit the dog?". Similarly, *Wang et al.* does not show, teach or suggest defining a set of conditions comprising a condition stipulating that, for a clause in the natural language text database to constitute an answer to a question clause, at least one of the constituents of the question clause should have a

corresponding constituent in the clause having the same syntactic function and an equivalent lexical meaning as claimed in claims 1 and 13.

Additionally, *Wang et al.* merely discloses using an interrogative pronoun to identify a semitype of interest, which is a result of analysis of the semantics of the question and is not the syntactics. Thus nothing in *Wang et al.* shows, teaches or suggests determining a syntactic function of interest based on an analyzing step and identified interrogative pronoun as claimed in claim 16. Rather, *Wang et al.* merely discloses using the interrogative pronoun to identify a semitype of interest.

Also, *Wang et al.* merely discloses using the semitype for matching. Thus, nothing in *Wang et al.* shows, teaches or suggests identifying a constituent having a predetermined syntactic function in the question clause or defining a set of conditions...a word token...being comprised in a constituent having the same syntactic function as the identified constituent in the question clause as claimed in claim 16. Applicant respectfully traverses the Examiner's statement in the Office Action that 'the SemType value person plays an important role,...where questions, location features are important'. Applicants respectfully point out that the SemType value person with respect to the "who" questions and location features with respect to "where" questions are not the syntactic functions but refer to the semantic types of the words.

Since the focus in *Wang et al.* is on the semantic type, Applicant respectfully submits that *Wang et al.* teaches away from the use of the syntactic function. Thus nothing in *Wang et al.* shows, teaches or suggests the same use of the syntactic function as claimed in claims 1, 13 and 16. Therefore, Applicant respectfully

requests the Examiner withdraws the rejection to claims 1, 13 and 16 under 35 U.S.C. §102(a).

Claims 2-4, 9-12 and 17-20 depend from claims 1, 13 and 16 and recite additional features. Applicant respectfully submits that claims 2-4, 9-12 and 17-20 would not have been anticipated by *Wang et al* within the meaning of 35 U.S.C. §102(a) at least for the reasons as set forth above. Therefore, Applicant respectfully requests the Examiner withdraws the rejection to claims 2-4, 9-12 and 17-20 under 35 U.S.C. §102(a).

Claims 5-8 were rejected under 35 U.S.C. §103 as being unpatentable over *Wang et al.* in view of *Hedin et al.* (U.S. Patent No. 5,386,556). Claims 14-15 were rejected under 35 U.S.C. §103 as being unpatentable over *Wang et al.* in view of well known prior art.

Applicant respectfully traverses the Examiner's rejection of the claims under 35 U.S.C. §103. The claims have been reviewed in light of the Office Action, and for reasons which will be set forth below, Applicant respectfully requests the Examiner withdraws the rejection to the claims and allows the claims to issue.

As discussed above, since nothing in *Wang et al.* shows, teaches or suggests the primary features as claimed in claims 1 and 13, Applicant respectfully submits that the combination of *Wang et al.* with the secondary reference to *Hedin et al.* or well known prior art will not overcome the deficiencies of the primary reference. Therefore, Applicant respectfully requests the Examiner withdraws the rejection to claims 5-8 and 14-15 under 35 U.S.C. §103.

The prior art of record, which is not relied upon, is acknowledged. The references taken singularly or in combination do not anticipate or make obvious the claimed invention.

Thus it now appears that the application is in condition for reconsideration and allowance. Reconsideration and allowance at an early date are respectfully requested.

If for any reason the Examiner feels that the application is not now in condition for allowance, the Examiner is requested to contact, by telephone, the Applicant's undersigned attorney at the indicated telephone number to arrange for an interview to expedite the disposition of this case.

In the event that this paper is not timely filed within the currently set shortened statutory period, Applicant respectfully petitions for an appropriate extension of time. The fees for such extension of time may be charged to our Deposit Account No. 02-4800.

In the event that any additional fees are due with this paper, please charge our Deposit Account No. 02-4800.

Respectfully submitted,

BURNS, DOANE, SWECKER & MATHIS, L.L.P.

Date: August 10, 2005

By: _____
Ellen Marcie Emas
Registration No. 32,131

P.O. Box 1404
Alexandria, Virginia 22313-1404
(703) 836-6620